The Tai Languages: Key to Austro-Tai[*]

Paul K. Benedict

Ormond Beach, Florida, U.S.A

Throughout the early decades of this century, among Eastern as well as Western scholars, a recurrent topic concerned the origin of the Thai and other Tai peoples, with language serving as a key indicator. Linguists of that period were inordinately fond of employing tone as a basic feature of classification, unaware of the fact that tone and even tonal systems themselves can be borrowed, as long demonstrated now in Southeast Asia. The great French scholar, Henri Maspero, was so enamoured of the tones of Vietnamese that he saw fit to classify that language with the Tai family despite the basically Mon-Khmer affiliation of the lexicon, including the numerals. In the case of the Tai languages the situation was discovered to be somewhat different; here the tones reflect a three-tone proto-system identical with that of Chinese, with lexical agreements in the numerals and a scattering of other items, notably the culture-bound trio: horse, saddle, and ride. A closely similar situation was found to exist in the Hmong-Mien (then Miao-Yao) family, with lexical agreements in the same lexical areas. Given the linguistic climate of the period, the linguistic conclusion here was predictable: the Tai and Hmong-Mien languages are to be classified together with Chinese and the related Tibeto-Burman languages under a single stock: Indo-Chinese.

When his first published paper appeared (Benedict, 1939), the writer was hardly acquainted with Thai or other Tai languages but subsequently came to study them in some detail, aided in no small measure by the comparative studies of Maspero. He had been trained at Harvard as an anthropologist and, somewhat later, had come under the influence of Morris Swadesh, who was attempting to measure linguistic relationships by agreements shown in a basic or "core" list of 100 items, with scores for related languages typically ranging at 20% or higher, at times down to 10% or so. Somewhat to his surprise, the writer found that the Chinese/Tai score is zero! Ironically, even one and two, the only numerals in the primitive-oriented Swadesh list, differ from the other, higher numerals in failing to show a Chinese-Tai correspondence. Clearly, it seemed to him, something is amiss in all this; if Chinese and Tai were truly related, as widely believed even by distinguished scholars at the time, they apparently would be the only pair of related languages in the world with a zero score.

Austro-Tai (AT) began with a search for a genuine rather than spurious relationship for the Tai family, and the Tai languages have continued to play a key role in uniting that widespread linguistic stock, including now Japanese and Hmong-Mien as well as Tai, along with other members of the Kadai superfamily, and Austronesian, a thousand languages or so, dispersed over a large part of the globe. At the very onset the Tai languages played a major role in providing the key forms for a

---

[*]**Editor's Note:** The author has asked that he be allowed to present an essay in which is distilled the essence of what he has discoveed about the Tai languages and the ancestry of the Tai peoples without the encumbrance of the vast scholarly apparatus that would be needed for proper documentation. For further details and sources see Benedict (1988), as well as his numerous articles in the journal *Kadai*.

two roots became more precise, in Austronesian as well as in Tai, the phonological correspondence became clarified, in a manner often encountered when working with related languages and in this case reducing the chance factor to a near-zero level. This merits examination in some detail:

eye   Proto-Malayo-Polynesian (PMP) ***mata**   P-Tai ***ta**^{A}

die   PMP ***m-atay**   P-Tai ***taay**^{A}

The PMP reconstructions of the time (early 1940s), based largely on the work of the German scholar Otto Dempwolff, excluded from consideration forms from the archaic Formosan languages, required for the later Proto-Austronesian (PAN) reconstructions. Similarly, the available P-Tai reconstructions did not have the benefit of forms from Saek, the archaic Northern Tai speech so magnificently recorded by William Gedney. On the basis of the roots as then reconstructed, however, it all seemed very simple, perhaps even too simple: the proto-language involved, later christened Proto-Austro-Tai (PAT), had roots of the form ***mata** 'eye' and ***m-atay** 'die' (contrasting with ***p-atay** 'kill'); the ancestral Tai languages, in the course of becoming monosyllabic, had simply elided the first syllable and assigned to the product tone *A, by far the most common (equal to *B +*C).

At that time, the age of innocence in Southeast Asian linguistics, this pair of roots presented only two "problems" in Tai: the Central Tai languages show /th/~/h/ reflexes in these two roots, rather than the anticipated /t/, and the long /aa/ in 'die' remains unexplained. These two "problems," as linguists like to call them, were to lead further afield, with a remarkable convergence of the pathways in AN and Tai. In the AN field, for the /t/ of these two roots the Formosan languages were found to show a series of highly idiosyncratic reflexes, symbolized by Dyen's cover *C, altogether distinct from the reflexes, usually /t/, for the standard PAN ***t**, indicating that something other than a simple ***t** must be reconstructed here at an earlier (PAT) level. On the mainland, within the same general time frame, it was discovered that Saek has the totally unanticipated /pr/ in these roots: Saek **pra**^{A} 'eye,' **praay**^{A} die.' It was now clear that the Central Tai /th/~/h/ reflects an earlier, P-Tai ***pr** (/r/>/h/ is common Tai shift); further, that a cluster such as ***pr** is also indicated at the PAT level, with Dyen's *C serving as the PAN reflex, since PAN lacks clusters of this type. The writer, who came upon the Formosan evidence first, was more than a little happy when the Saek forms made their appearance, making for this remarkable convergence of pathways.

The Tai languages have continued to present a treasure-trove of linguistic riches. The unexplained /aa/ of P-Tai ***praay**^{A} 'die' has been found to reflect the /a/ of the first syllable of the PAT root ***m-apray** (> PAN ***m-aCay**). This feature, which the writer has called "vocalic transfer" (VT), furnishes the key to an understanding of Gedney's "puzzles" and other vocalic alternations in Tai, often yielding doublet forms through variable VT, e.g., in the "core" AT root for 'I,' where Malay has /aku/, Southwest Tai (SWT) varies between /ku^{A}/ (Thai, Lao), without VT, and /kaw^{A}/ (Ahom, Shan, Khamti), reflecting VT, while Central Tai (CT) has /kaw^{A}/ and Northern Tai (NT) has /ku^{A}/; in another "core" AT root, for 'fire,' where Malay has /apuy/, reflected in Kam (Kam-Sui group) **pui**^{A}, Thai **fay**^{A} reflects the /a/ of the initial syllable, paralleled elsewhere in SWT and in CT but with NT reflecting only final **-i**, without VT. In his justly celebrated *Handbook of Comparative Tai*, the late Li

Fangkuei made a valiant attempt at constructing a P-Tai vocalic system without recourse to VT. David Strecker has shown, however, that a parallel process underlies complex doublet formation in the closely related Kam-Sui group, making it virtually impossible to extend the Li line of reconstruction to the parent Proto-Kam-Tai because of the vast number of often weird rimes that such a procedure would entail.

The Tai languages, in addition to displaying VT, also play a valuable role in reflecting PAT-level functors of various kinds. Let us examine the PAT "core of core" kinship term root for 'child.' The perfect cognate pair here, Malay **ana?** and Thai **luuk**, understandably failed to make it even through the initial sorting of MP and Tai forms, which demanded something more than a final **-k** in common to be retained. Convergence then set in, as often seen in the study of related forms. In the AN field, the Formosan cognates that were uncovered revealed that the Malay (and PMP) medial **-n-** reflects a PAN-level ***-l-** (regular shift), making the root PAN ***alak**. In the Kadai field, meanwhile, Li was able to reconstruct Proto-Kam-Sui (PKS) ***laak** 'child' from ***alak** through VT (see above). We are now halfway to the Thai (and general SWT/CT) /luuk/ but badly in need of help with the vocalism. Further AN studies, notably by Dyen, attested to the presence at times of an initial labial element (Dyen's ***W-**) in the AN root. This has now been identified as a variable ***u-** nominal marker, further modifying the root to PAN ***:(u-)alak**. Japanese, which retained this functor until a fairly late stage, as shown by its use in loans from Chinese (**uma** 'horse,' et al.), parallels AN in its cognate here **wara-** 'child' from ***u-la[k]-** (regular shifts). The source of the Tai vocalism is now evident: ***u-lak .> *uluk** (through assimilation) > **luuk** (through VT); Tai reflects the ***u-** marker while Kam-Sui does not, making for a doublet development somewhat akin to that seen above in vocalism.

The Tai languages are also of special value as a linguistic tie, so to speak, providing for important connections at early AT levels. The writer, with an anthropological speciality in kinship terminology, feels particularly indebted to Tai and other Kadai languages for their help in this regard. The ancestral Austro-Tai proto-language appears to have inverted the widespread /pa/ 'father' and /ma/ 'mother' but AN and Tai has each retained only onehalf the evidence: PAN ***ama** 'father' (***a-** found in other basic kinship terms); P-Tai ***pa**^{B} (Li's ***pa**^{C}) 'aunt, parent's older sister'; PKS ***pa**^{B} 'aunt, parent's older sister,' reflecting an extension of the maternal role, as shown by Lakkia (closely related to Kam-Sui) **pa**^{B} 'mother.' This precious Tai/Kadai evidence ties in with the Japanese (Jp.): **ama** 'grandfather,', a dialectal extension from an earlier 'father'; **haha**, Old Jp. **FaFa~Fa** < ***pa(pa)** 'mother'; Jp ***oba**,, Old Jp. **woba** (typical medial voicing) 'little mother' ='aunt,', paralleling the Tai/Kadai development.

Austro-Tai, now massively documented, has far-reaching implications for the culture history of Southeast Asia. The homeland was in China, long before the ancestral Chinese entered the northwest corner of that country in the upper reaches of the Yellow River. The product of its speakers was rice culture, first developed, it now appears, in the fertile middle Yangtse valley, eventually spreading along with other cultural traits to the ancestral Chinese to the north, followed by extensive cultural interchange on all sides. The complex loanword material still being sorted out, at different levels, includes substantial evidence for the relatively late spread from Chinese to Kadai and and Hmong-Mien of numerals, used in trade, and the horse complex (see above). The Hmong-Mien were home-bodies, moving down into Southeast Asia only much later, while the ancestral Austronesians and Japanese moved to the coast and finally out to sea, and the ancestral Kadai peoples moved early and

in large numbers far to the south (Thailand) and the southwest (Shan), even reaching Assam (the extinct Ahom). We do now, at long last, have a reasonably precise answer to the old question: who are the Tai and where did they come from?

REFERENCES

Benedict, P.K. (1939). Semantic differentiation in Indo-Chinese. *Harvard Journal of Asiatic Studies, 4*, 213-229.

Benedict, P.K. (1988). Kadai linguistics: The rules of engagement. In J.A. Edmondson & D.S. Solnit (Eds.), *Comparative Kadai: Linguistic studies beyond Tai* (pp. 323-338). Dallas: SIL/VTA Series in Linguistics, Publication 86.